**Tugas 2**

Nama : Patricia Joanne

NPM : 140810160065

**3 April 2019**

Metode Elbow

**Program:**

function [IDX,C,SUMD,K] = best\_kmeans(X)

% [IDX,C,SUMD,K] = best\_kmeans(X) partitions the points in the N-by-P data matrix X

% into K clusters. Rows of X correspond to points, columns correspond to variables.

% IDX containing the cluster indices of each point.

% C is the K cluster centroids locations in the K-by-P matrix C.

% SUMD are sums of point-to-centroid distances in the 1-by-K vector.

% K is the number of cluster centriods determined using ELBOW method.

% After find the best K clusters, IDX, C, SUMD are determined using kmeans function in matlab.

dim=size(X);

% default number of test to get minimun under differnent random centriods

test\_num=10;

distortion=zeros(dim(1),1);

for k\_temp=1:dim(1)

[~,~,sumd]=kmeans(X,k\_temp,'emptyaction','drop');

destortion\_temp=sum(sumd);

% try differnet tests to find minimun disortion under k\_temp clusters

for test\_count=2:test\_num

[~,~,sumd]=kmeans(X,k\_temp,'emptyaction','drop');

destortion\_temp=min(destortion\_temp,sum(sumd));

end

distortion(k\_temp,1)=destortion\_temp;

end

variance=distortion(1:end-1)-distortion(2:end);

distortion\_percent=cumsum(variance)/(distortion(1)-distortion(end));

plot(distortion\_percent,'b\*--');

[r,~]=find(distortion\_percent>0.9);

K = r(1,1)+1;

[IDX,C,SUMD] = kmeans(X,K);

**Screenshot:**

